

PERBAIKAN METODE B.GATOS UNTUK RESTORASI CITRA DOKUMEN KUNO NON-LINIER



Arliansyah

J2A 604 006

Skripsi

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains

Pada Jurusan Matematika

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS DIPONEGORO

SEMARANG

2009

ABSTRAK

Dalam pemeliharaan sebuah dokumen kuno, saat ini dokumen kuno sudah mulai disimpan dalam format digital, karena dokumen dalam format digital tidak akan berubah sepanjang waktu. Tugas akhir ini menyajikan teknik untuk memperbaiki citra dokumen kuno non-linier berdasarkan pendeteksian kata dan baris kata yang diambil dengan menggunakan *scanner*. Teknik yang digunakan diaplikasikan pada citra yang memiliki tingkat keabuan. *Pre-Processing* pada citra dokumen kuno berfungsi untuk mendapatkan citra biner. Konsep awal pendeteksian objek tekstual pada citra dokumen kuno dengan menggunakan *connected component*. Rotasi dan shift dilakukan dengan menggunakan acuan garis upper dan garis lower pada tiap kata yang terbentuk dengan bantuan *box hand*. Tahap terakhir adalah restorasi secara keseluruhan dari citra asli dokumen kuno non-linier yang ditentukan oleh hasil pelurusan citra biner. Hasil akhir dari citra dokumen kuno adalah citra dokumen kuno yang sempurna dan telah dikembalikan dalam bentuk *grayscale*.

Kata Kunci:

Pre-Processing, connected component, rotasi dan shift, box hand, restorasi final.

ABSTRACT

In ancient document maintenance, an ancient document has been saved in digital format. It is because of document which is saved in digital format will not change all along time. This study provided a technique to improving an ancient non-linear document images based on discovery words dan line words which is taken by scanner. This technique is applicated to grayscale images. Pre-Processing on ancient document image has a function to get binary images. The first concept of textual object detection ancient document image is useing a connected component. Rotation and shift are useing with upperline and lowerline hint on each word that has been formed with box hand. last phase is restoration as a whole from ancient document original of non-linear determined from result of streamlining of binary image. The final result of ancient document images are ancient document image that is perfect and back to grayscale form.

Key words : Pre-Processing, connected component, rotation and shift, box hand, final restoration.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring dengan perkembangan dan kemajuan di bidang teknologi informasi membuat segala sesuatu mulai beralih ke peralatan yang menggunakan sistem komputerisasi. Informasi tidak hanya disajikan dalam bentuk teks, tetapi juga dapat berupa gambar, audio (bunyi, suara, musik), dan video.

Dokumen dalam format digital memiliki beberapa kelebihan dibandingkan dengan dokumen kertas. Dokumen digital lebih mudah disebarluaskan sehingga dapat dimanfaatkan oleh banyak orang. Selain itu, penelusuran teks juga lebih efektif bila dilakukan pada dokumen digital. Dokumen kuno merupakan suatu dokumen yang usianya sudah tua dan kebanyakan sudah mengalami kerusakan akibat termakan usia. Permasalahan yang muncul adalah dokumen-dokumen tersebut pada proses digitalisasinya tidak selalu dalam keadaan baik. Beberapa dokumen ada yang memiliki orientasi teks melengkung akibat proses *scan* yang tidak sempurna. Dokumen kuno merupakan salah satu peninggalan sejarah yang harus mulai dilestarikan dalam bentuk digital. Selain bisa digunakan oleh banyak orang, dokumen kuno dalam bentuk digital juga tidak akan berubah sepanjang waktu.

Dengan berbagai kondisi di atas, perlu adanya restorasi (pembangunan kembali) agar citra dokumen kuno dapat lebih mudah

dibaca. Restorasi meliputi penghilangan *noise* dan *smoothing*, pengembalian orientasi teks melengkung menjadi mendatar serta meluruskan orientasi *background* agar mengikuti citra teks yang telah lurus. Pada saat ini sudah banyak teknik untuk melakukan restorasi citra salah satunya adalah teknik yang di ajukan oleh B.Gatos, yaitu untuk meluruskan dan memperbaiki orientasi teks yang melengkung digunakan bantuan garis lengkung yang didapat dengan menggunakan persamaan regresi linier dan menggunakan *box hand* untuk melakukan pendeteksian baris kata dan kata.

Tugas akhir ini membahas bagaimana memperbaiki teknik yang diajukan B.Gatos untuk mendapatkan citra dokumen kuno dalam bentuk digital, pendeteksian baris kata, dan meluruskan baris kata yang melengkung, serta perancangan antarmuka agar perangkat lunak yang telah jadi lebih mudah untuk dioperasikan. Tugas akhir ini diimplementasikan dalam pembuatan perangkat lunak simulasi restorasi citra hasil *scan* dokumen kuno. Simulasi dibuat menggunakan pemrograman berbasis GUI (*Graphical User Interface*) yang tersedia pada MATLAB 7.0.1

1.2. PERMASALAHAN

Permasalahan yang muncul dalam pembuatan aplikasi ini antara lain sebagai berikut:

- 1) Bagaimana memperbaiki teknik merestorasi citra dokumen kuno non-linier yang diajukan oleh B.Gatos dengan menggunakan citra biner sebagai acuan dan meluruskan baris kata pada bagian citra dokumen kuno yang melengkung dengan mengubah orientasi kata agar mendatar.
- 2) Bagaimana teknik pembuatan perangkat lunak untuk merestorasi citra dokumen kuno dengan menggunakan MATLAB 7.0.1

1.3. PEMBATASAN MASALAH

Berdasarkan uraian di atas, maka batasan masalah pada tugas akhir ini adalah:

- 1) Restorasi dilakukan pada citra hasil *scan* dokumen kuno dengan warna *grayscale*
- 2) Sudut rotasi $-90^{\circ} < \alpha < 90^{\circ}$
- 3) Sebuah citra hanya memuat satu halaman dokumen kuno.
- 4) Dokumen kuno merupakan hasil tulisan tangan yang diambil melalui *scanner* dan hanya berisi teks, tidak berisi gambar.
- 5) Format citra input berbentuk jpg, jpeg, dan bmp.
- 6) Pengujian hasil program dilakukan dengan pengamatan secara visual (pengamatan mata).
- 7) *Tool* yang digunakan untuk pengerjaan tugas akhir ini adalah MATLAB 7.0.1.

1.4. TUJUAN PENULISAN

Tujuan umum dari penulisan tugas akhir ini adalah membuat perangkat lunak yang dapat merestorasi citra hasil *scan* dokumen kuno yang melengkung sehingga bagian teks pada citra yang melengkung menjadi lurus dan memiliki orientasi mendatar dengan tetap mempertahankan nilai artistik dari dokumen tersebut.

1.5. SISTEMATIKA PENULISAN

Sistematika pembahasan dalam tugas akhir ini terbagi menjadi 4 bab yang dimulai dari bab pendahuluan dan terakhir adalah bab penutup.

BAB I PENDAHULUAN berisi tentang latar belakang masalah, permasalahan, pembatasan masalah, tujuan penulisan serta sistematika penulisan.

BAB II DASAR TEORI berisi tentang dasar teori *Pre-Processing* yaitu pembentukan citra digital, deteksi kata dan baris kata, pelurusan teks, pengembalian citra ke dalam *grayscale*, serta MATLAB 7.0.1

BAB III PEMBAHASAN berisi tentang bahan dan alat yang digunakan dalam restorasi citra dokumen kuno, algoritma restorasi citra dokumen kuno yang dibagi menjadi 4 bagian utama, perancangan program simulasi, penggunaan program simulasi, hasil dan pembahasan.

BAB IV PENUTUP berisi kesimpulan dari pembahasan tentang restorasi citra dokumen kuno.